IMPROVED ILLUMINATION-INVARIANT DESCRIPTORS FOR ROBUST COLOUR OBJECT RECOGNITION

R.J. O'Callaghan and D.R. Bull

Image Communications Group,
Centre for Communications Research, University of Bristol,
Woodland Road, Bristol, BS8 1UB, U.K.

ABSTRACT

Colour object recognition is heavily influenced by variation in the scene illumination conditions. This paper proposes a set of illumination-invariant descriptors of image content. The descriptors are based on a moment-based approach to histogram comparison and, in the case of an object imaged under two different lighting conditions, permit straightforward recovery of the illumination change involved. The efficacy of the descriptors is compared experimentally with a variety of existing techniques, using an established methodology and an existing purpose-built dataset. The evidence suggests that the new descriptors outperform existing techniques in the area of colour object recognition.

1. INTRODUCTION

Colour is widely held to be one of the most powerful descriptors of image content. As a result it has become an almost ubiquitous component in content-based retrieval systems, as well as many other computer vision applications. It has, in various implementations, been shown to be relatively robust to background complication and can be independent of image/object size and orientation [1]. However, without dedicated pre-processing, colour features suffer substantial degradation due to the effects of varying illumination. This places severe limitations on the direct use of colour as a descriptor and much research has addressed the problem of generating illumination-independent, yet reliable colour features.

It is important to differentiate between two alternative approaches to illumination invariance. Firstly, there is the theory of "colour constancy", in which colour pixel (rgb) responses (obtained under unknown illumination) are mapped to illumination independent descriptors. These are most commonly the predicted appearances of the pixels under a known "canonical" illuminant. The philosophy behind this approach stems from experiments in human vision, which have demonstrated that the perception of object colour is, to a large degree, independent of variations in scene illumination. In computer vision, therefore, colour constancy pre-processing predicts the image that would have been captured were the scene to have been lit by the canonical illuminant. For object recognition, this adjusted image becomes the input to a conventional colour-based algorithm, such as Swain and Ballard's colour indexing [2], or similar.

An alternative approach is to compute a set of illumination invariant statistics or properties directly from the image, but without predicting the appearance of the object under a canonical illuminant [3], [4], [5]. The use of invariants is an inherently simpler approach to object recognition, because, unlike colour constancy, no attempt is made to recover information about the unknown illuminant. This also means that fewer constraining assumptions are necessary in the algorithm. The key difference between this and colour constancy is that the illumination invariant features are now extracted directly from the image. In this paper we demonstrate the power of this latter approach for object recognition.

2. COLOUR IMAGES

Colour image formation depends, in general, on three factors, as expressed by equation (2.1). The value registered by the kth sensor at location X is well approximated by:

$$\rho_k^X = \int e(\lambda) s^X(\lambda) f_k(\lambda)\, d\lambda \ , \qquad 1 \le k \le n, \quad (2.1)$$

where λ denotes wavelength, $e(\lambda)$ is the spectral power distribution of the incident light, $s^X(\lambda)$ is the spectral reflectance of the object at location X and $f_k(\lambda)$ is the spectral response of the kth sensor. For standard colour imaging, there are three sensors (n=3), corresponding to

0-7803-7402-9/02/$17.00 ©2002 IEEE

the red, green and blue channels. The integral of equation 2.1 thus amounts to a projection from an infinite dimensional space into three dimensions. In this model, the relationship between recorded pixel values (i.e. the 3-vector, $\boldsymbol{\rho}$) under a change in illumination is given by a simple matrix transform:

$$\boldsymbol{\rho}_{\text{illum1}} = \mathbf{A}\,\boldsymbol{\rho}_{\text{illum2}} \qquad (2.2)$$

where $\mathbf{A}$ is, in general, a 3x3 matrix. This results in the colour histogram of the image undergoing an affine transformation in the three dimensional rgb space:

$$H_{\text{illum1}}(\boldsymbol{\rho}) = H_{\text{illum2}}(\mathbf{A}\,\boldsymbol{\rho}) \qquad (2.3)$$

In this work, we will constrain $\mathbf{A}$ to be a diagonal matrix and derive new illumination invariants based on this model. The use of the diagonal transformation, also known as Von Kries Adaptation or coefficient rule, has a long history in colour constancy [6][7].

3. PROPOSED ALGORITHM

The problem of colour object recognition is now posed as a shape recognition exercise in histogram space. In order to develop illumination independent descriptors for the images, we must attempt to devise transformation-invariant shape descriptors in histogram space. In [8], Taubin and Cooper describe such a moment-based approach to object recognition. Their geometric invariants were applied by Healey and Slater in histogram space to achieve illumination invariance.

3.1 Centred-Moment Matrices

We will now briefly review Taubin and Cooper's centred-moment method, referring the reader to [4] for a more detailed presentation or to [8] for a full treatment of the theory. The matrix of monomials of degree two is defined as:

$$\mathbf{X}_{[1,1]}(\boldsymbol{\rho}) = \mathbf{X}_{[1,1]}(\rho_1, \rho_2, \rho_3) = \begin{pmatrix} \rho_1^2 & \rho_1\rho_2 & \rho_1\rho_3 \\ \rho_1\rho_2 & \rho_2^2 & \rho_2\rho_3 \\ \rho_1\rho_3 & \rho_2\rho_3 & \rho_3^2 \end{pmatrix} \qquad (3.1)$$

To arrive at the associated centred-moment matrix, this matrix of monomials is integrated over histogram space as follows:

$$\mathbf{M}_{[1,1]} = \frac{1}{|H(\boldsymbol{\rho})|} \int \mathbf{X}_{[1,1]}(\boldsymbol{\rho} - \bar{\boldsymbol{\rho}})\, H(\boldsymbol{\rho})\, d\boldsymbol{\rho} \qquad (3.2)$$

$$\text{where} \quad \bar{\boldsymbol{\rho}} = \frac{1}{|H(\boldsymbol{\rho})|} \int \boldsymbol{\rho}\, H(\boldsymbol{\rho})\, d\boldsymbol{\rho}$$

$$|H(\boldsymbol{\rho})| = \int H(\boldsymbol{\rho})\, d\boldsymbol{\rho}$$

3.2 Diagonal Invariants

Healey and Slater [4], having made the assumption that the histograms are related by a general 3x3 transformation, used a set of affine invariant descriptors in their recognition system. Here, however, we have made a different, diagonal, assumption. So, instead of affine invariants, we require properties that are only invariant to diagonal transformations of the histogram.

Any affine transformation can be realised as the product of a change of coordinates followed by an orthogonal transformation. The matrix, $\mathbf{L}$, for the coordinate transformation is given by the Cholesky factorisation of the second order centred moment matrix [8]:

$$\mathbf{L}\,\mathbf{M}_{[1,1]}\,\mathbf{L}^T = \mathbf{I} \qquad (3.3)$$

In general, the intermediate histograms are related by an orthogonal transformation, $\mathbf{Q}$:

$$H(\mathbf{L}\boldsymbol{\rho}) = \tilde{H}(\mathbf{Q}\tilde{\mathbf{L}}\boldsymbol{\rho}) \qquad (3.4)$$

However, using the diagonal assumption, it can be shown that $\mathbf{Q}$ must be the identity matrix. Giving:

$$H(\mathbf{L}\boldsymbol{\rho}) = \tilde{H}(\tilde{\mathbf{L}}\boldsymbol{\rho}) \qquad (3.5)$$

In other words, when transformed by the relevant $\mathbf{L}$ matrices, the (centred) histograms are equal. This also allows straightforward recovery of the transformation matrix:

$$\mathbf{A} = \mathbf{L}^{-1}\tilde{\mathbf{L}} \qquad (3.6)$$

For our experiments we have chosen two sets of statistics as recognition features. Firstly, it was noted that the transformation matrix retrieved by equation 3.6 will only be correct if the diagonal assumption holds. If the two histograms are not related by a diagonal transformation, the matrix, $\mathbf{A}$, returned will be lower triangular instead of diagonal. Therefore, we might use the norm of the off-diagonal elements as a distance measure. To derive a symmetrical function, we also repeat the calculation, having swapped $\mathbf{L}$ and $\tilde{\mathbf{L}}$ in the matrix multiplication. The sum of the two results is then used as the first part of our distance measure. Secondly, we calculate the 6x6 matrix $\mathbf{M}_{[2,2]}$ for each of the $\mathbf{L}$-transformed histograms. As the matrix elements are not independent we need only store a vector of 15 distinct values. The second distance measure we use is given simply by the Euclidean distance between these vectors, for the respective histograms.

3.3 Specularity Elimination

The assumptions of a Mondrian world are not satisfied in real images. As a final refinement to the proposed algorithm, we applied a colour normalisation as a pre-processing step, in an attempt to make the system more robust to confounding phenomena, such as specular reflection. The normalisation applied is given by

$$\frac{N/_3\, r}{\sum_{i=1}^{N} r_i}, \frac{N/_3\, g}{\sum_{i=1}^{N} g_i}, \frac{N/_3\, b}{\sum_{i=1}^{N} b_i} \qquad (3.7)$$

where N is the number of pixels in the image and r, g and b are the pixel values of the red, green and blues channels respectively. This normalisation is one of two used by Finlayson *et al.* [9] in a standalone algorithm. We use it here to reduce the impact of specularities on the moment-invariants.

4. EXPERIMENTS

In order to evaluate the algorithm critically in terms of object recognition performance, we have recreated the experiments described in Funt *et al.* [10] and Barnard *et al.* [11]. The original datasets for these experiments are publicly available at http://www.cs.sfu.ca/~colour/data/. See [12] for details.

In [10] and [11] the authors compare the abilities of a range of colour constancy algorithms in enhancing colour object recognition. The two papers use different datasets, but otherwise the paradigm is broadly similar. The flowchart of figure 1 illustrates the experimental procedure. The images are of objects under a number of different illuminants. For each test, one illuminant is chosen as the canonical, giving a subset of "model" images. Each of the test images (the remainder of the dataset) is then corrected, using colour constancy, to correspond to the chosen canonical illuminant, after which similarity matching is performed, using histogram intersection [2]. The results are presented in terms of percentages of matches in a given rank over a number of tests. A "rank one" match occurs when the correct target image (i.e. the model image portraying the correct object) is rated the most similar to the test image. For a "rank two" match, the correct image would be returned as the second most similar model image, and so on.

The proposed algorithm compares very favourably with the original results for colour constancy algorithms as well as subsequently reported results for

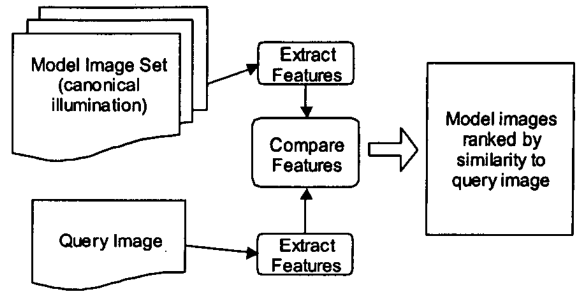

***Figure 1.** Object recognition experiment*

illumination invariant recognition [3]. Table 1 shows the results on the first dataset [13], while table 2 shows those for the second, larger set [14]. "Nothing" refers to the results of straightforward object recognition without colour constancy pre-processing. "Actual" refers to the use of the measured rgb values of each illuminant to directly calculate the diagonal illumination change matrix. For completeness we have also included a set of results from an implementation of Healey and Slater's method [4], applied to the first dataset.

For both experiments, we have referenced the best performing colour constancy algorithms from the authors' original work. For the first experiment, these are the white-patch retinex and 3D gamut-constraint methods [10]. In the second experiment they are two variations of 3D gamut mapping algorithms (ND-CRULE-AVE, ECRULE-SCWIA-12) [11]. The multimodal neighbourhood signature (MNS) is a subsequently published method for illumination invariant recognition, which uses invariant features extracted from localised regions [3]. Koubaroulis *et al.* have reported its performance on the first of the two datasets and we have included those results in Table 1 for further comparison.

In both experiments, the diagonal moment invariants outperform all the other algorithms in terms of rank one matches. There is also favourable comparison between our results and those for true colour constancy, where the real illuminant is used to correct the images. The improved performance figures for the second experiment may be partly attributed to the inclusion of "zero-distance" queries in the results. The procedure differs subtly in this latter experiment, in that the model images are also included in the test dataset. A colour constancy algorithm, unaware that the images are identical, inevitably creates error by attempting to "correct" the query image. Direct illumination invariant features, on the other hand, will always find a trivial perfect match with the target in this case. Since the task of same-illumination recognition is undoubtedly a valid one, this is a systematic advantage of illumination invariant statistics over colour constancy pre-processing.

Also noteworthy is the amount of data necessary to support each algorithm. In the algorithms using colour

	Rank 1	Rank 2	Rank 3
Actual	92.3	5.5	1.4
Nothing	28.4	10.2	9.1
Diagonal Invariants	81.8	6.8	2.7
"L" matrix only	68.2	14.5	7.3
Retinex	67.7	8.6	5.9
3D Gamut	67.3	8.6	5.5
MNS	60	15	7
Moment Invariants [4]	46.4	19.5	12.3

Table 1. *First Dataset: % correct matches, by rank*

	Rank 1	Rank 2	Rank 3
Actual	87.7	10.6	1.2
Nothing	42.3	14.0	6.0
Diagonal Invariants	85.9	4.1	1.0
ND-CRULE-AVE	80.9	10.4	1.5
ECRULE-SCWIA-12	72.7	12.4	3.9

Table 2. *Second Dataset: % correct matches, by rank.*

constancy, for instance, two-dimensional chromaticity histograms are stored for both model and test images. In the reported implementation, these are 16x16, requiring storage of 256 values. To obtain the results of Tables 1 and 2, we have calculated just 21 values per image. Perhaps even more impressively, the performance of the "**L**" measure alone (68% rank one) requires storage of only 6 values. This still exhibits performance similar to the colour constancy methods.

5. CONCLUSION

We have presented an algorithm for the generation of moment-based illumination invariants of images. In the context of object recognition, these descriptors are robust to changes in position and orientation and, given prior segmentation of the region of interest from the background, are independent of object size. Further, they do not suffer from the limitations of colour constancy pre-processing in the situation of same-illumination recognition. Experimental evidence shows that a small number of these statistics offer very favourable performance in comparison to existing illumination-invariant methods for object recognition. Tests were performed following an established experimental routine, using a publicly available database of images. These results show that directly computed invariant features can offer a more viable solution to the object recognition problem than current colour constancy methods.

ACKNOWLEDGEMENTS

This work was supported by EPSRC grant GR/M84183, under the Link project Autoarch.

REFERENCES

[1] Y. Rui, T.S. Huang and S.F. Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues", Journal of Visual Communication and Image Representation, Vol. 10, pp 39-62, March 1999.

[2] M.J. Swain and D.H. Ballard, "Color Indexing", International Journal of Computer Vision, Vol. 7, No. 1, pp 11-32, 1991.

[3] D. Koubaroulis, J. Matas and J. Kittler, "Illumination Invariant Object Recognition Using the MNS Method", Proceedings of the 10th European Signal Processing Conference, Tampere, Finland, pp 2173-2176, Sept. 4-8, 2000.

[4] G. Healey and D. Slater, "Global color constancy: recognition of objects by use of illumination invariant properties of color distributions", Journal of the Optical Society of America A, Vol. 11, No. 11, pp 3003-3010, Nov. 1994.

[5] G. Healey and D. Slater, "Computing Illumination-Invariant Descriptors of Spatially Filtered Color Image Regions", IEEE Transactions on Image Processing, Vol.6, No. 7, pp 1002-1013, July 1997.

[6] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Diagonal Transforms Suffice For Color Constancy", Proceedings of the 4th International Conference on Computer Vision (ICCV'93), Berlin, Germany, pp 164-171, May 11-14, 1993.

[7] K. Barnard, "Computational color constancy: taking theory into practice", M.Sc. Thesis, Simon Fraser University, 1995.

[8] G. Taubin and D.B. Cooper, "Object Recognition Based on Moment (or Algebraic) Invariants", in J. Mundy and A. Zisserman, eds., *Geometric Invariance in Computer Vision*, MIT Press, Cambridge, Mass., pp 375-397, 1992.

[9] G.D. Finlayson, B. Schiele, and J.L. Crowley, "Comprehensive colour image normalization", In Proceedings European Conference on Computer Vision (ECCV'98), Freiburg, Germany, pp 475-490, June 2-6 1998.

[10] B. Funt, K. Barnard and L. Martin, "Is colour constancy good enough?" Proceedings of the 5th European Conference on Computer Vision (ECCV'98), Freiburg, Germany, pp 445-459, June 2-6, 1998.

[11] K. Barnard, B. Funt and L. Martin, "Colour Constancy Meets Colour Indexing", Submitted for publication, Oct. 9, 2000. (http://www.cs.berkeley.edu/~kobus/research/publications/ indexing/)

[12] K. Barnard, L. Martin, B. Funt, and A. Coath, "Data for Colour Research", Color Research and Application, in press. (www.cs.berkeley.edu/~kobus/research/publications/data_for_colour_research/)

[13] http://www.cs.sfu.ca/~colour/image_db/index.html.

[14] http://www.cs.sfu.ca/~colour/data/objects_under_different_lights/index.html.